To appear in Proc. Symposium on Interactive 3D Graphics, Snowbird, Utah, March 1990.

A Real-time Optical 3D Tracker For Head-mounted Display Systems

Jih-fang Wang
Vernon Chi
Henry Fuchs

Department of Computer Science
University of North Carolina
Chapel Hill, NC 27599-3175

1989

Abstract

In this paper, a new optical system for real-time, three-dimensional position tracking is described. This system adopts an "inside-out" tracking paradigm. The working environment is a room where the ceiling is lined with a regular pattern of infrared LEDs flashing under the system's control. Three cameras are mounted on a helmet which the user wears. Each camera uses a lateral effect photodiode as the recording surface. The 2D positions of the LED images inside the field of view of the cameras are detected and reported in real time. The measured 2D image positions and the known 3D positions of the LEDs are used to compute the position and orientation of the camera assembly in space.

We have designed an iterative algorithm to estimate the 3D position of the camera assembly in space. The algorithm is a generalized version of the Church's method, and allows for multiple cameras with nonconvergent nodal points. Several equations are formulated to predict the system's error analytically. The requirements of accuracy, speed, adequate working volume, light weight and small size of the tracker are also addressed.

A prototype was designed and built to demonstrate the integration and coordination of all essential components of the new tracker. This prototype uses off-the-shelf components and can be easily duplicated. Our results indicate that the new system significantly out-performs other existing systems. The new tracker provides more than 200 updates per second, registers 0.1-degree rotational movements and 2-millimeter translational movements, and processes a working volume about 1,000 ft^3 (10 ft on each side).

Introduction

Significant advance has been made towards realistic synthesis and display of three-dimensional objects using computers during the last two decades. It is now a common practice to generate complicated 3D scenes with hidden surfaces removed, and visible surfaces realistically lighted and smoothly shaded. However, the interaction between human and computer-generated scenes remains largely remote and two dimensional through devices such as mice, joysticks, keyboards, etc.

An ideal interactive mechanism was proposed by Ivan Sutherland [Sut65] in 1965. The concept of the *Ultimate Display* is that the host computer controls the existence of objects in a virtual world. Computer-generated chairs could be sat upon, and computer-generated bullets would be fatal. In this system, users experience and interact with computer-generated objects just as they do with real physical objects. Unfortunately, the Ultimate Display concept is a goal far from attainable even with today's technology. However, we are now in a position to investigate certain feasible subsets of the Ultimate Display concept. The head-mounted display represents such a subset.

For more than a decade, research effort has been expended to develop head-mounted display systems for easy interaction with computers. In a head-mounted display system, the computer-generated images are presented on the screens of two small video displays mounted in front of the user's eyes. As the user moves, his or her head position is constantly measured, and appropriate views of a computer-generated 3D environment are displayed to provide the illusion of a virtual world. For example, the head-mounted display system developed at UNC has been used to allow a user to "walk-around" in a virtual environment.

The walk-around concept, while not as powerful and realistic as the one advocated in the Ultimate Display, does provide a natural and effective human/machine interaction mechanism. In general, the head-mounted display provides a more realistic mechanism for visualizing and interacting with virtual 3D objects than conventional displays. It allows viewpoint selection through natural head and body

movements. It also provides a means of interacting with computer-generated objects using the pose of the user's hand. Users of the system learn complex 3D structures and relationship with less effort than if conventional 2' interaction devices are used. Since the method harnesses the naturally well-trained hand-eye-body coordination, it should prove to be a more effective paradigm for human-machine interaction.

A head-mounted display system consists of three major components: a graphics engine, a 3D position tracking subsystem, and a helmet-mounted display. The head-mounted system developed at UNC uses the Pixel-Planes machine [FGH+85], which is capable of generating thirty thousand Gouraud shaded polygons per second, as the graphics engine. The helmet is equipped with two color-liquid-crystal television sets, each with a two-inch-diagonal display screen. The user's head position is tracked by a Polhemus 3D position tracker. The Polhemus tracker consists of a source radiating magnetic waves which are picked up by a sensor attached to the helmet. The sensor detects the magnetic field generated by the source to infer the position of the user's head.

Much improvement needs to be made before this head-mounted display system can be used as a practical human/computer interaction device. In this research, we concentrate on improving the performance of the tracking subsystem. The Polhemus tracker has a slow update rate and suffers the lag (latency) problem. The lag, which is the time between the user's movement to the output from the Polhemus changes accordingly to report the movement, can be as long as 120 milliseconds [CHB+89]. One should note that the problems of the update rate and latency are different. For example, one can have a tracker with fast update rate but the report still lags behind the head motion. The Polhemus also has limited working range (the user has to stay within a small distance from the radiating source). Furthermore, the function of the Polhemus is affected by magnetic perturbations in the environment. In a laboratory where radiating sources (e.g. TV sets, computers and terminals) and metallic surfaces abound, the performance of the Polhemus can be seriously degraded. Hence, the goal of this research is to design and construct a new tracker which provides a large working range, high accuracy in the estimated head position, fast update rate, low latency, and immunity to the electro-magnetic interference of the environment.

The remainder of this paper is organized as follows: We first survey the state-of-the-art in motion tracking and discuss the fundamental working principles of the new tracker. Algorithms for recovering 3D position are then presented, and the system's error is analyzed. A prototype system was constructed to prove the correctness of the new concept. The performance of this prototype was quantitatively measured and is presented in this paper.

Background

Although our primary application of 3D tracking systems is in head-mounted displays, such tracking devices have also found applications in interactive surface design and 3D modeling, and have been used as unconstrained 3D graphic s input tools. Below we briefly survey the existing 3D tracking devices.

Commercial and experimental 3D position tracking devices have used acoustic, magnetic, mechanical, and optical methods for reporting 3D position. Acoustic ranging systems use the time-of-flight principle to estimate the range of objects in space. Because the speed of sound varies if ambient air density changes, these systems have poor accuracy over a large range. Also, acoustic systems can not sense orientation directly.

The Polhemus 3D position tracker [Pol80] is a magnetic system consisting of a sensor which detects a low-frequency magnetic field generated by a source. The performance of the Polhemus is affected by any conducting materials present in the environment. Further, the Polhemus has a limited working range ($\sim 1\ m^3$), and a update rate ($\sim$ 16 updates/sec) [CHB+89] which is barely enough for interactive applications.

The first mechanical linkage head-mounted display system was first built at the University of Utah [Sut68][Vic74]. The Argonne Remote-Manipulator (ARM) at UNC [Kil76] and the Noll box [Nol71] also fall into this category. These types of systems have a limited working range. Besides, the friction inertia of the systems and the mechanical linkage attached to the user greatly restrict the motion of the user. It is also difficult to track several objects simultaneously with these kinds of systems.

Many commercial and experimental trackers use optical sensors. The optical tracking method is appealing because it is relatively insensitive to environmentally induced distortion, has a large working volume, and can be made fast and accurate. Below we briefly survey some commercially available optical trackers.

SELSPOT [Wol74][LO74] and OP-EYE [Uni81] are two commercial systems consisting of camera-like units using lateral effect photodiodes as the sensing surfaces. These systems detect a single light source focussed on the photodiode and determine the 2D location where the light beam strikes the photodiode surface. A pair of these cameras can be used to estimate the 3D location of the light source by stereoptic means. The SELSPOT system does not report 3D position in real time. The OP-EYE system has poor resolution and a very limited working volume.

OPTOTRAK [Nor88] uses one camera with two dual-axis CCD infrared position sensors. Each position sensor has a dedicated processor board to calculate the image

position of the light source. Again, the triangulation principle is applied to recover the position of the light source in space. The system is expensive, and the bulky camera weighs more than 10 pounds.

Gary Bishop [Bis84][BF84] proposes a new scheme which uses several 1D sensors mounted on the helmet to observe the environment. By clustering sensors with different orientation together and pooling information on the image shift from all sensors, 3D movement of the helmet can be derived by solving a set of non-linear equations. Custom VLSI circuitry is built to enable the sensors to report the shift in the observed image pattern in real time.

The New Tracking Method

Head-mounted display systems require that the position of the user's head be tracked in real time with high accuracy in a large working environment. It is evident from the above discussion that none of the currently available systems surveyed has a satisfactory performance. However, the optical tracking method seems to possess the most potential. We propose to develop a new improved tracker using this method.

Most commercial optical tracking systems place the sensors at fixed locations. These sensors are used to observe light emitted from the light sources attached to the helmet which the user wears. Such schemes are termed *outside-in* tracking. Outside-in tracking methods, although intuitively simple and appealing, fail to provide the accuracy we want. For example, suppose several LEDs are affixed on antennae mounted on the helmet the user wears, with base line separation 0.5m. A 0.1° of rotation by the user moved the LEDs by only about 0.4mm. Moreover, to cover a large working volume in the outside-in tracking scheme, the cameras must have wide fields of view. The required resolution of these cameras is not feasible in existing technology.

Our system is similar to the one proposed by Gary Bishop. In our system, we reverse this outside-in configuration. Instead, we affix many light sources, or beacons, on the ceiling of the room and mount the camera on the helmet the user wears. Since many beacons are used in this *inside-out* configuration, the camera needs to cover only a small field of view. Also, a small amount of rotational movement induces much larger shift of the LED images on the photodetector surface. Hence to detect the same 0.1° rotation, a much less resolution is needed.

To achieve a high update rate, our system uses a lateral effect photodiode as the sensing surface of the camera. A lateral effect photodiode features a large photosensitive surface, usually square in shape (1 cm^2), on which the x and y location of the centroid of incident luminous flux is measured in nearly real time (~ 200 μsec). Using a lateral-effect photodiode as the sensing surface has the following

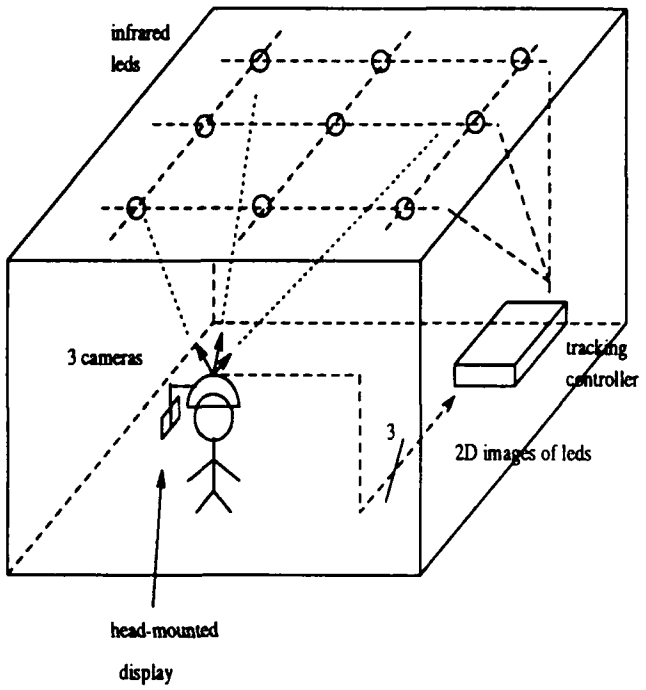

Figure 1: Configuration of the proposed tracking system

advantages over conventional CCD arrays: a lateral-effect photodiode provides faster response and higher positional resolution, has no dead-zone over the sensing surface, and provides an accurate positional reading even if the image is somewhat out of focus (or blurred). Finally, there is no need to compute the centroid of the detected light spot.

The requirement of a large working volume is met by using many light emitting diodes (LEDs) with high output power and a wide emission angle as beacons [Sei85]. The room is lined with a regular pattern of LEDs on the ceiling, and these LEDs are flashing under the system's control. The LEDs radiate in the near infrared so that the user is not distracted by their constant blinking. Figure 1 depicts the configuration of the proposed tracking system.

Algorithms for Inferring 3D position

For an outside-in tracking scheme, the 3D position of a light source is routinely computed using triangulation where at least two cameras are used. Each camera determines a line along which the light source must lie. The 3D position of the light source is then located by intersecting at least two such projection lines in space.

Position recovery is a more complex issue in an inside-out configuration. The problem can be formulated as follow: given the position $A_i[a_{i1}, a_{i2}, a_{i3}, 1]$ in the homogeneous world coordinate of several beacons, and the image locations $B_i[b_{i1}, b_{i2}, 1]$ of the beacons under a perspective projection, the goal is to find a transformation matrix $M_{4\times3}$—which represents the projection process—such that M satisfies the following constraint:

$$[a_{i1}, a_{i2}, a_{i3}, 1]M = [b_{i1}, b_{i2}, 1]. \quad (1)$$

Because there are twelve unknowns in M, and each observed beacon provides two independent constraints based on Equation 1, at least six beacons are needed to solve a set of twelve linear equations to derive M.

Since a photodiode is used as the sensing surface, only one beacon can be turned on inside the field of view of the camera at any time. Hence, beacons are flashed sequentially. During the period of time between the flashing of the first beacon and the flashing of the sixth beacon, the user might have moved. This movement will introduce error in the recovered 3D position of the camera. Thus, we need an algorithm which can compute the 3D position of the camera with the least number of beacons.

A method proposed by Earl Church [Chu45] was first used in aerial photogrammetry. Church's method determines the position of the camera from an aerial photograph by locating three known landmarks in the photograph. It can be shown that since only three beacons are used, there is no closed-form solution and an iterative method has to be employed to derive the 3D position of the camera.

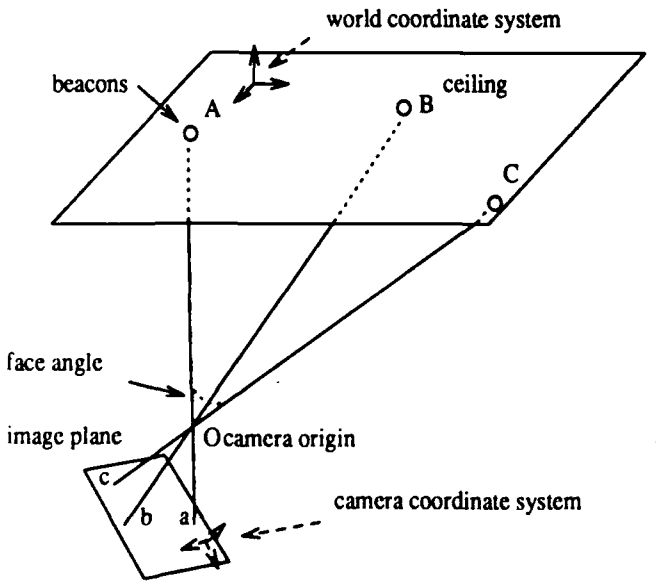

Figure 2: Church's algorithm

Figure 2 shows how Church's method can be applied in our tracking environment. Two coordinate systems are defined here. The world coordinate system is defined on the fixed room environment, while the camera coordinate system is defined by using the nodal point of the camera as origin. In Church's method, three beacons are observed by the camera. Any two of the three observed beacons form a *face angle* with the camera origin. Church's solution is based on the condition that the face angle subtended by any two beacons in space is equal to the face angle subtended at their corresponding image locations. The face angles formed by the camera origin and the image projection of the beacons can be calculated directly in the camera coordinate system. However, because the location of the camera nodal point in the world coordinate system is unknown, the face angles subtended by the beacons in space are unspecified.

Church's method starts by guessing a value for the camera origin in the world coordinate system. With this hypothesized camera origin position, we can form three face angles in the world coordinate system. These face angles will in general not match those computed from the images unless the hypothesized camera position is correct. By partial differentiation of the difference between the two sets of face angles with respect to the current hypothesized position (X_h, Y_h, Z_h), we can get an adjustment $(\Delta X_h, \Delta Y_h, \Delta Z_h)$ which is added back to the hypothesized position. The process is iterated until the adjustment becomes insignificant, and the hypothesized position converges to the correct position.

Our simulation results indicate that Church's algorithm is quite robust even in the presence of input errors (discussed later). Although Church's method requires iterations to converge to the correct solution, the convergence can be expedited if the initial guess is close to the true location of the camera. In our application, we can always use the last head position as an educated initial guess.

System Error Analysis

In this section, we discuss the errors in the system and how they affect the accuracy in the reported 3D position. To quantify the error of the system, we assume that Church's method, when given perfect input data, converges to the correct 3D position. Then the error in the computed 3D location must be due to input errors propagating to the output [PW83]. Hence, errors in Church's method come from the measurement errors in the two sets of face angles used as inputs to the algorithm. The measurement error in the face angles formed by the camera origin and the beacon images is mainly due to the limited resolution of the photodiode and the nonlinear characteristics of the photodiode surface. While the measurement error in the face angles formed by the camera origin and beacons in space is mainly due to the positional uncertainty of the beacons.

First, we study the effect of the limited photodiode

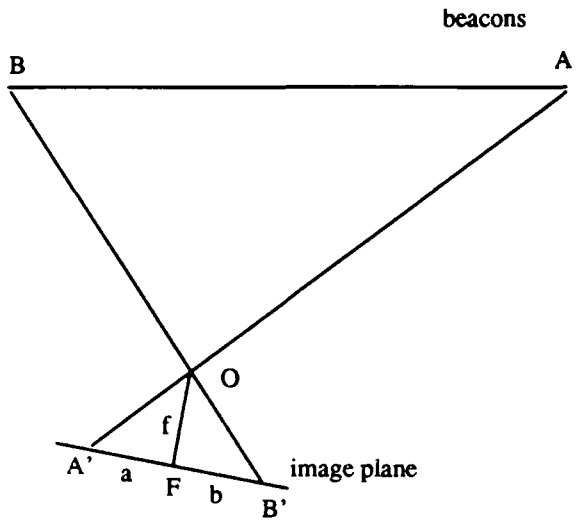

Figure 3: Error in the image plane

resolution on the accuracy of the face angles measured in the image. Figure 3 shows the plane formed by two beacons A and B and the nodal point of the camera O. The image positions of beacons A and B are denoted by A' and B', respectively. It can be seen that[1]

$$\begin{aligned}\theta &= \angle A'OB' = \angle A'OF + \angle FOB' \\ &= \tan^{-1}\frac{a}{f} + \tan^{-1}\frac{b}{f},\end{aligned}$$

where f is the focal length of the camera lens. From [PW83], we have

$$\begin{aligned}\varepsilon_\theta &= \frac{\partial\theta}{\partial a}\varepsilon_a + \frac{\partial\theta}{\partial b}\varepsilon_b + \frac{\partial\theta}{\partial f}\varepsilon_f \\ &= \frac{f}{a^2+f^2}\varepsilon_a + \frac{f}{b^2+f^2}\varepsilon_b + \left(\frac{a}{a^2+f^2} + \frac{b}{b^2+f^2}\right)\varepsilon_f,\end{aligned}$$

where ε_θ is the measurement error of the face angle θ, and $\varepsilon_a, \varepsilon_b$ and ε_f represent the absolute error bound on the measurement of a, b, and f. The maximum error in the measurement of a and b, by definition, is one half the smallest measurable unit of the photodiode. If the area of the photodiode is D^2 and the resolution of the photodiode is r, then $\varepsilon_a = \varepsilon_b = D/2r$. The error in measuring the focal length is a constant and is independent of different readings from the photodiode. We ignore this error term for now. Thus we have:

$$\varepsilon_\theta = \left(\frac{1}{a^2+f^2} + \frac{1}{b^2+f^2}\right)\frac{fD}{2r}. \qquad (2)$$

Equation 2 states that in order to minimize the measurement error of the face angles in the camera coordinate system, a lens with a long focal length should be used, the separation between image points should be as far apart as possible, and the resolution of the photodiode should be as high as possible.

The error in the face angles measured in the world coordinate system is computed again using Figure 3. If the coordinates of A and B are (x_a, y_a), and (x_b, y_b), then

$$\angle AOB = \cos^{-1}\frac{\overrightarrow{OA}\cdot\overrightarrow{OB}}{|OA||OB|} = \cos^{-1}\frac{x_a x_b + y_a y_b}{\sqrt{x_a^2+y_a^2}\sqrt{x_b^2+y_b^2}}.$$

Let the maximum error in the position of beacons be ε_p. then $\varepsilon_x = \varepsilon_y = \varepsilon_p$. So we have

$$\begin{aligned}\varepsilon_\theta &= \left(\frac{\partial\theta}{\partial x_a} + \frac{\partial\theta}{\partial y_a} + \frac{\partial\theta}{\partial x_b} + \frac{\partial\theta}{\partial y_b}\right)\varepsilon_p \\ &= \left(\frac{x_a+y_a}{|OA|^2} + \frac{x_b+y_b}{|OB|^2}\right)\varepsilon_p. \qquad (3)\end{aligned}$$

We see that this error is directly proportional to that in the placement of beacons. Again a large separation between beacons decreases the error.

Single View vs. Multiple Views

We found that the above requirements impose conflicting demands on the system design. For example, a lens with a long focal length has a relatively small field of view. As discussed earlier, with a small viewing field, we obtain a high translational sensitivity with less resolution. However, in order to locate at least three beacons inside this small viewing field, beacons must be placed close together, which implies a lot of them are needed. How to measure their positions accurately becomes difficult. Furthermore, closely-placed beacons introduce larger errors in the face angle measurements in the world coordinate system, and limit the separation of their images on the photodetector.

In order to achieve large face angles without sacrificing the accuracy and resolution of the tracking system, more than one view are needed. If multiple views with wide angular separation are used to observe different parts of the environment, each view needs to cover only a small area with at least one beacon inside. Each view can then provide the necessary high accuracy in its 2D position reading. The large angular separation of the views induces wide image separation and therefore large face angles for the combined assembly. The advantage of this multiple-view concept is that small translational movements are readily detected due to the small area each view covers, while widely separated views induce large baseline separation among beacons. These mutually conflicting design criteria can thus be optimized independently instead of interdependently in this configuration.

[1] The formulas derived in this section are for a restricted case where errors are only two dimensional. That is, we consider only error in the plane formed by camera origin O, and beacons A and B. Formulas for general cases can be found in [Wan89].

Implementing Multiple Views

In realizing the multiple-view concept, three cameras and lenses with long focal length are used. These cameras have wide angular separation and each camera covers only a small field of view.

To recover the 3D position of the whole camera unit is a more difficult problem than if a single camera is used. With a single camera, light from different beacons always passes through the same nodal point and focuses on a single image plane. The nodal point of the lens is thus naturally chosen to be the origin of the camera coordinate system, and the face angles in both world and camera coordinate systems are readily calculated. If multiple cameras are used and each camera has its own nodal point, light from different beacons focuses onto different image planes. In Figure 4, we show two cameras with two distinct nodal points O_a and O_b observing beacons A and B respectively. These two cameras are separated by a fixed distance. By carefully measuring the camera positions, the vector $\overrightarrow{O_bO_a}$ in the camera coordinate system can be derived. Hence, the face angle in the image coordinate system can be calculated by translating one of the nodal point (e.g. O_b) to merge with the other (O_a). In order to derive the corresponding face angle in space, we need to know the expression of $\overrightarrow{O_bO_a}$ in the world coordinate system. However, the direction of $\overrightarrow{O_bO_a}$ in the world coordinate depends on the current head position which is unknown. Hence, Church's method fails when there are multiple cameras with distinct nodal points.

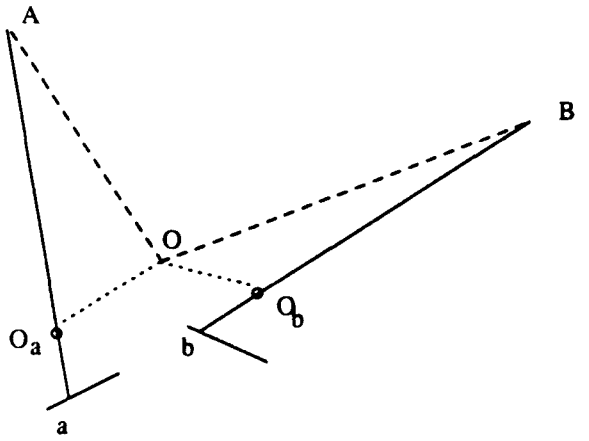

Figure 4: Multiple cameras with separated nodal points

We now present a solution which generalizes Church's method for this multiple camera configuration. In the generalized method, the origin of the camera coordinate system O is arbitrarily chosen, referring to Figure 4. Now following Church's method, let's hypothesize the position of the virtual origin O in world coordinates. We can compute the face angle $\angle AOB$ in the world coordinate system in terms of the hypothesized position of O. The face angle in the camera coordinate system is derived as follow: $\overrightarrow{OO_a}$ can be measured in the camera coordinate system so can $\overrightarrow{O_aa}$. We can thus derive the expression of the angle $\angle OO_aa$ as

$$\angle OO_aa = \cos^{-1}\left(\frac{\overrightarrow{O_aa}\cdot\overrightarrow{O_aO}}{|\overrightarrow{O_aa}|\ |\overrightarrow{O_aO}|}\right).$$

The same for $\angle OO_bb$. Using the law of cosines,

$$|OA|^2 = |O_aO|^2 + |O_aA|^2 - 2\,|O_aA|\,|O_aO|\cos(\angle AO_aO),$$

we compute $\overrightarrow{O_aA}$, and similarly, $\overrightarrow{O_bB}$. Since $\overrightarrow{OA} = \overrightarrow{OO_a} + \overrightarrow{O_aA}$, and $\overrightarrow{O_aA} = -\overrightarrow{O_aa}\,|O_aA|\,/\,|O_aa|$, we can derive $\overrightarrow{OA}$ in the camera coordinate system; similarly for $\overrightarrow{OB}$. Then the face angle $\angle AOB$ in the camera coordinate system can be computed. Using the two sets of face angles as defined above with respect to the virtual camera origin O, Church's method again becomes applicable.

System Performance Evaluation

A prototype system was constructed to prove the correctness of our design. The prototype consists of three identical photodiode cameras mounted on a helmet (Figure 5). The helmet is positioned on a mounting device which provides six degrees of freedom in motion (translation along the x, y and z directions and rotation about the x, y and z axes) that can be precisely measured. Special signal processing circuitry was constructed to filter the data from the cameras. A parallel interface bridges the output from the circuitry to a μVax-II workstation which runs the generalized Church's algorithm for 3D position recovery. The calculated position of the camera assembly is displayed using the Pixel-Planes machine. This desktop prototype is shown in Figure 6. We quantitatively measured the speed, range, and accuracy of the prototype. Results are summerized below.

Speed

The tracking process consists of two distinct phases: ***sampling*** and ***computing***. The sampling phase starts when the host sends control signals to flash three beacons and initiate the signal processing circuitry, and ends when data are acquired from the circuitry. The computing phase starts immediately afterwards to calculate the 3D position and orientation of the camera assembly using the generalized Church's algorithm.

Our experiments on a μVax-II workstation indicate that the time spent on sampling is much smaller than that

machine	speed (updates/sec)
Sun 3/50	< 1
Sun 3/60	7
Sun 4	82
μVax-II	25
μVax-3200	69
DECstation-3100	215

Table 1: Speed comparison on different machines

spent on computing. The pure sampling rate of the system, without doing any position computation, can be as high as 1500 Hz. If complete cycles of sampling and computing are performed, the rate drops to about 25 updates per second on a μVax-II.

Although the update rate of this prototype is only comparable to that of most commercial trackers, the performance of the system can be improved significantly using a fast host machine. We estimated the update rate of the tracker—using several different host computers—by running the generalized Church's algorithm on them. Note that since the sampling circuitry is hard-wired to the μVax-II, the time spent on the sampling phase for these hosts cannot be estimated this way. We just use the figure (1500 Hz) from μVax-II as an estimation for all the host computers. We ran the same algorithm for position estimation on several Sun workstations: 3/50, 3/60, and Sun 4, and several DEC workstations: μVax 3200 and DECstation 3100. Table 1 shows the average update rate on different machines. From the table, we conclude that it is possible to achieve a speed for near real-time performance if a fast host machine is used. With a fast host computer, we can also cut down the lag to about 5 milliseconds.

Range and Accuracy

To estimate the working range, one has to note that range and accuracy are tightly related. Accuracy depends on the resolution of the photodiode, which in turn depends on the strength of the light signal. As the working range is made larger by moving the light sources farther away from the cameras, the received light energy decreases. This decrease in the signal-to-noise ratio degrades the resolution and hence the accuracy.

Our goal here is to achieve an accuracy of less than 1cm in position error and about 0.3 degree in orientation error in a room with about 4m on each side. The maximal tolerable error corresponds roughly to a shift of 2 pixels in the image—assuming that the user is looking at a target 2 meters away with a 90 degree field of view, and an image resolution of 512 by 512 pixels. Our calculation shows that a photodiode resolution of at least 1 part in 1000 is needed. Based on this, we conducted experiments to estimate the working range of the system. An infrared LED was mounted on the pen holder of an x-y plotter, so its movement can be controlled by a host computer. A camera was placed 3m away from the plotter surface. In our experiment, the LED traversed a straight line, and the positions of the LED after each 0.5mm movement were recorded. This movement corresponded roughly to a 1 part in 1200 resolution on the photodiode surface. Figure 7 shows the LED locations reported by the tracker. The curve shows good linearity. The same experiment conducted at 3.5m still shows good linearity but with slightly more jitter in the output. The results demonstrate that the prototype has at least a 3m working range.

We also conducted experiments to estimate the accuracy of the prototype. Translational accuracy was measured by moving the camera assembly on the mounting stage through 1mm increments along one of the axes, while rotational accuracy was measured by rotating the assembly through 0.1° increments about one of the rotational axes. Figures 8 and 9 depict the results from our experiments.

From these figures, it can be seen that the prototype can register 0.1° rotational and 2 mm translational movements. This extremely high sensitivity in detecting both the rotational and translational motions can be attributed to the use of the inside-out tracking paradigm and the multiple-view concept. Our results clearly demonstrate the superiority of the new design.

Future Research

The next step is to construct a full working system in a 26x12x9 ft^3 room, using about one thousand infrared LEDs. The LEDs will be affixed to 2x2 ft^2 panels in 4 by 4 grids. These panels will be installed as ceiling tiles in the room. We are currently designing circuit boards with power and ground lines laid out in a rectangular grid. Beacons will be affixed at the grid junctions and can be easily addressed by enabling appropriate power and ground lines. We hope to have the fully working system by 1990.

Although the prototype successfully solves the problem of speed, range, and accuracy, the issue of light weight and small size remains to be addressed. The helmet weighs about 1 kg (each camera weighs 138 gram and each lens weighs 181 gram). This weight might cause fatigue after extended wear. We are currently seeking other technologies to reduce the weight of the tracker. One promising technology is holographic optics. According to [Tei88], it is possible to make a holographic lens the size of a silver dollar which can view multiple directions, focusing them onto a single photodetector. This new technology will trim down the weight of the current system by 10 fold.

Conclusion

This paper presents a new design concept for a 3D position tracking device. The new tracker has a large working volume, provides fast updates on the 3D position with low latency, and possesses better accuracy and resolution than currently available systems. The new tracker adopts an inside-out tracking method, with several widely separated views. A prototype was designed and built using off-the-shelf components for easy duplication, and its performance was quantitatively measured. The prototype demonstrates the feasibility of our design, and shows that the new tracker out-performs most commercially available devices. We expect this design to greatly enhance the usefulness of head-mounted display systems.

Acknowledgements

The first author wishes to thank Professor Frederick P. Brooks Jr. for his guidance and support. Thanks are also due to Dr. Gary Bishop of the Sun Microsystems and Dr. John Eyles for their helpful discussion, and to Mr. Brad Bennett and Mr. John Thomas of the Microelectronics Systems Laboratories of UNC for their help with system programming and hardware. Mr. Ron Azuma, who is now working on building the full system, provides helpful comments and criticisms. The work was supported in part by Office of Naval Research grant N00014-86-K-0680 and NIH Division of Research Resources grant RR 02170-05.

References

[BF84] T. G. Bishop and H. Fuchs. The self-tracker: A smart optical sensor on silicon. In *Proceedings Conference on Advanced Research in VLSI*. MIT Press, 1984.

[Bis84] T. G. Bishop. *Self-Tracker: A Smart Optical Sensor on Silicon*. PhD thesis, U. of North Carolina, Chapel Hill, NC, 1984.

[CHB+89] J. C. Chung, M. R. Harris, F. P. Brooks Jr., H. Fuchs, M. T. Kelley, J. W. Hughes, M. Ouh-Young, C. Cheung, R. L. Holloway, and M. Pique. Exploring virtual worlds with head-mounted displays. In *Proceedings SPIE Conference, Nonholographic True Three-Dimensional Display Technologies*, Los Angeles, CA, Jan. 1989.

[Chu45] E. Church. Revised geometry of the aerial photograph. In *Bulletin of Aerial Photogrammetry*, volume 15. Syracuse University, 1945.

[FGH+85] H. J. Fuchs, J. Goldfeather, J. P. Hultquist, S. Spach, J. Austin, F. P. Brooks Jr., J. Eyles, and J. Poulton. Fast spheres, textures, transparencies, and image enhancements in pixel-planes. *Computer Graphics*, 19(3), 1985.

[Kil76] P. J. Kilpatrick. *The Use of a Kinesthetic Supplement in an Interactive Graphics System*. PhD thesis, U. of North Carolina, Chapel Hill, NC, 1976.

[LO74] L. E. Lindholm and K. E. T. Oeberg. An optoelectronic instrument for remote on-line movement monitoring. *Bioelemetry*, 1, 1974.

[Nol71] M. A. Noll. *Man-machine Tactile Communication*. PhD thesis, Polytechnic Institute of Brooklyn, Brooklyn, NY 1971.

[Nor88] Northern Digital. *Trade literature on Optotrak - Northern Digital's Three Dimensional Optical Motion Tracking and Analysis System*. Northern Digital Inc., Waterloo, Ontario, Canada, 1988.

[Pol80] Polhemus Navigation Sciences Division. *3Space Isotrak User's Manual*. McDonell Douglas Electronics Company, 1980.

[PW83] S. M. Pizer and V. L. Wallace. *To Compute Numerically*. Little, Brown, and Company, Boston, MA, 1983.

[Sei85] Seimens. *Trade Literature on GaAlAs infrared emitters*. Seimens, Cupertino, CA, 1985.

[Sut65] I. E. Sutherland. The ultimate display. In *Proceedings IFIP Congress*, 1965.

[Sut68] I. E. Sutherland. A head-mounted three dimensional display. In *FJCC Conference Proceedings*, 1968.

[Tei88] M. Teitel, 1988. Personal communication with Mike Teitel from VPL.

[Uni81] United Detector Technology. *Trade literature on OP-EYE optical position indicator*. United Detector Technology Inc., Santa Monica, CA, 1981.

[Vic74] D. L. Vickers. *Sorcerer's Apprentice: Head-mounted Display and Wand*. PhD thesis, U. of Utah, Salt Lake City, UT, 1974.

[Wan89] J. F. Wang. *A Self-calibrating Real-time 3D Position Tracker*. PhD thesis, U. of North Carolina, Chapel Hill, NC, in preparation.

[Wol74] H. J. Woltring. New possibilities for human motion studies by real-time light spot position measurement. *Bioelemetry*, 1, 1974.

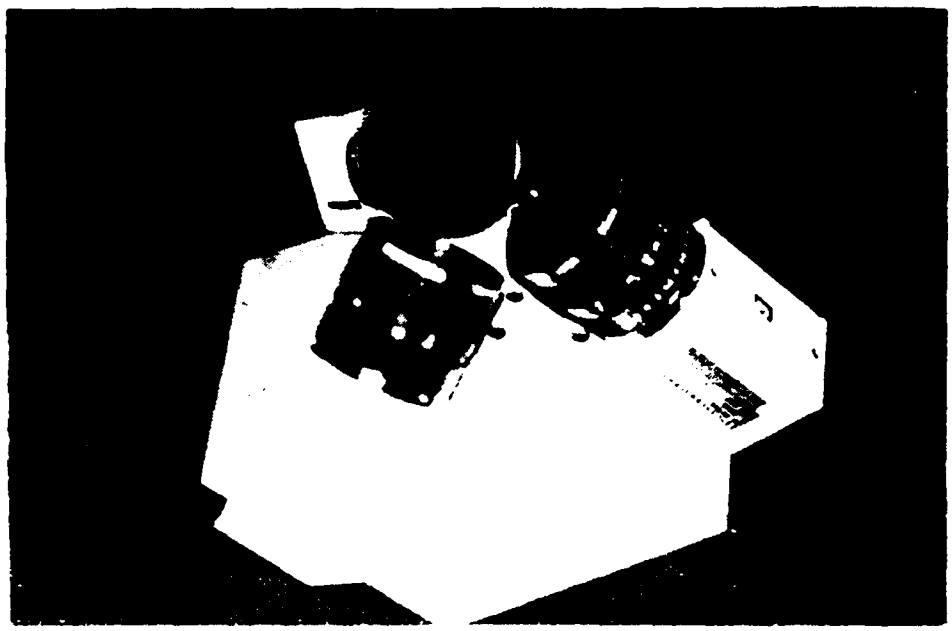

Figure 5: Three photodiode cameras mounted on a helmet

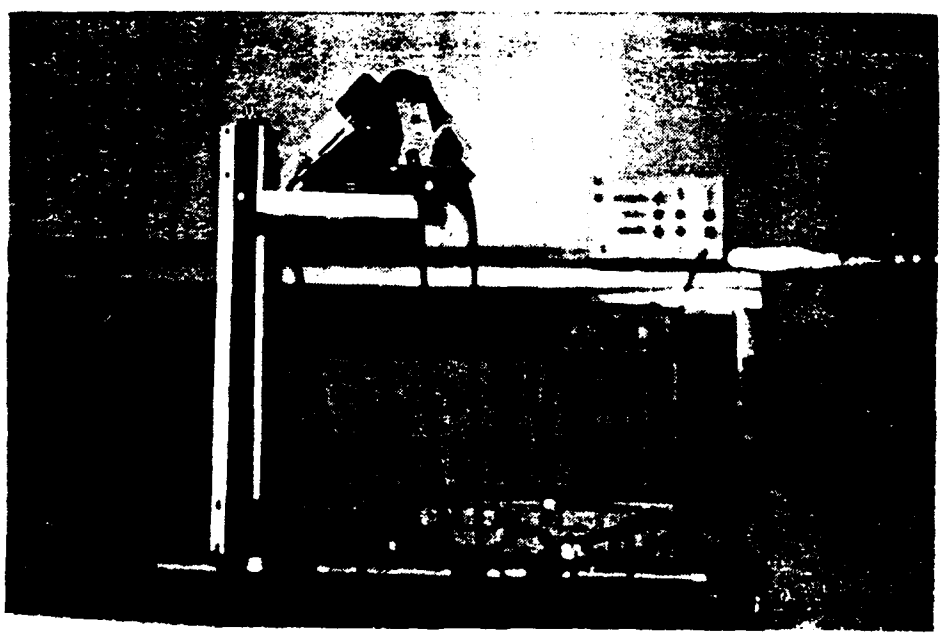

Figure 6: The desktop prototype

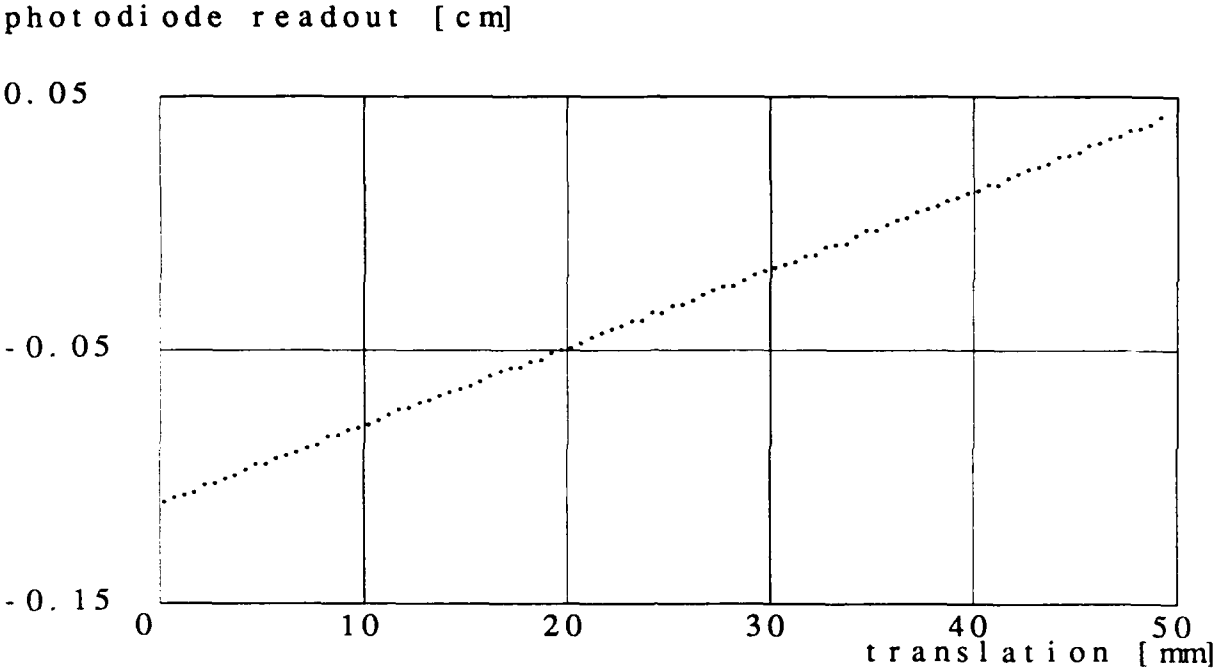

Figure 7: Photodiode resolution

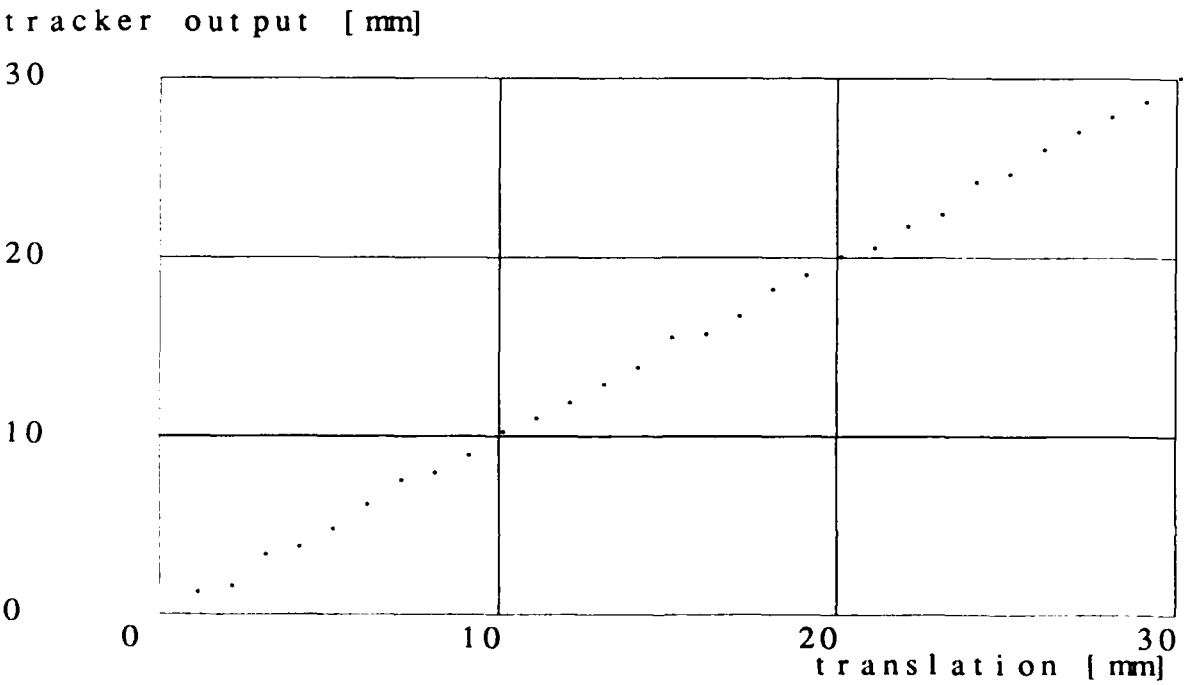

Figure 8: Translational sensitivity of the prototype

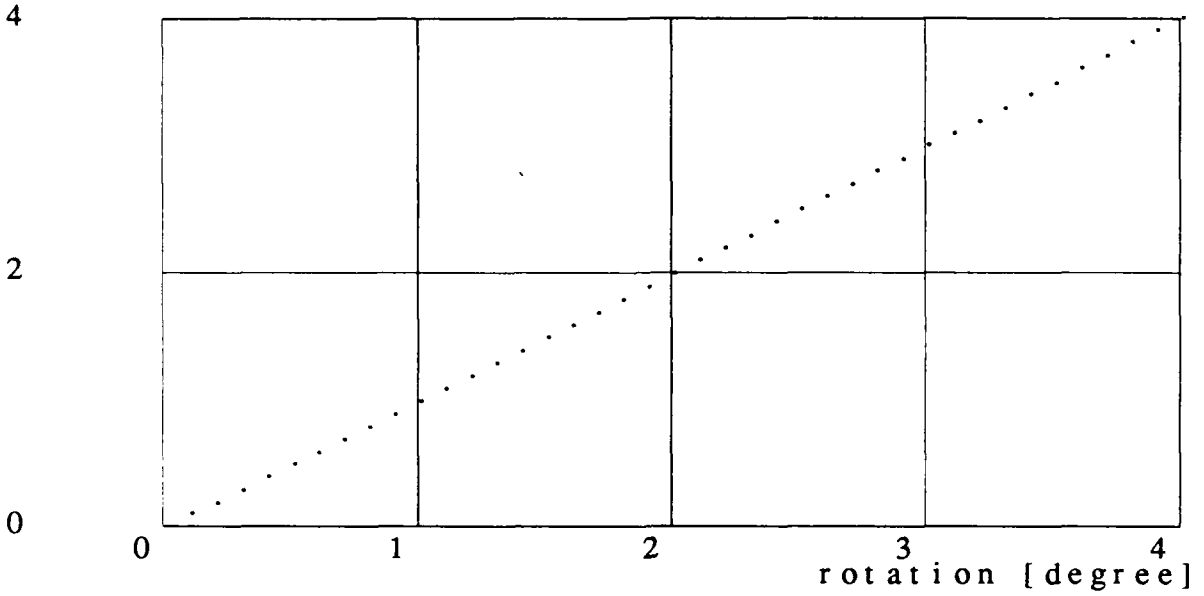

Figure 9: Rotational sensitivity of the prototype